R E P O R T   R E S U M E S

ED 011 938   24   AA 000 055

RESPONSE COMPLEXITY AND EXPERIMENTAL DESIGN.

BY- SNOW, RICHARD E.

REPORT NUMBER BR-5-0252-4   PUB DATE 18 NOV 66
CONTRACT OEC-6-10-078
EDRS PRICE MF-$0.09 HC-$0.52 13P.

DESCRIPTORS- *EDUCATIONAL RESEARCH, *RESEARCH DESIGN, *RESPONSE MODE, STATISTICAL ANALYSIS, APTITUDE TESTS, ACHIEVEMENT RATING, *MEASUREMENT TECHNIQUES, *LEARNING PROCESSES, STANFORD

THE SELECTION AND USE OF RESPONSE VARIABLES IN EDUCATIONAL EXPERIMENTS USING MULTIVARIATE ANALYSIS WERE CONSIDERED AND ASSESSED. RESPONSE VARIABLES WERE DEEMED CENTRAL CONCERNS FOR EITHER THEORETICAL OR PRACTICALLY ORIENTED RESEARCH, AND THEIR COMPLEXITY WAS DEALT WITH UNDER THE HEADINGS OF APTITUDE INPUT MEASURES, REPEATED LEARNING MEASURES, MULTIPLE LEARNING MEASURES, AND APTITUDE OUTPUT MEASURES. IT WAS CONCLUDED THAT, IF INSTRUCTIONAL METHODS AND PROCESSES ARE TO BE UNDERSTOOD AND IMPROVED, RESPONSE COMPLEXITY PROBLEMS MUST BE SOLVED. EARLY EMPHASIS SHOULD BE PLACED ON THE ASSESSMENT OF (1) THE NATURE OF STUDENT APTITUDES AS THEY INTERACT WITH TEACHING AND LEARNING PROCESSES, (2) THE COURSE OR PATTERNING OF THESE PROCESSES ACROSS THE OCCASIONS ON WHICH THEY OCCUR, (3) THE EXTENSITY OF INSTRUCTIONAL EFFECTS AS WELL AS THE INTENSITY OF ANY ONE EFFECT, AND (4) THE ENDURING CHANGES AND SUBSEQUENT EFFECTS OF LEARNING RELATIVE TO THE PATTERN OF INTELLECTUAL DEVELOPMENT IN GENERAL. THIS PAPER WAS PRESENTED AT THE ANNUAL STATE CONFERENCE ON EDUCATIONAL RESEARCH (18TH, SAN FRANCISCO, NOVEMBER 18, 1966). (GD)

BR-5-0252-4
OEC-6-10-078
P.A. 24

AA000055

Response Complexity and Experimental Design

Richard E. Snow
Center for Research and Development in Teaching
Stanford University

Presented at the 18th Annual State Conference on Educational Research
San Francisco, November 18, 1966

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education and Welfare, Office of Education, under the provisions of the cooperative research program.

The phenomena of classroom teaching and learning are obviously complex, so complex in fact that it is unreasonable to expect any one study to deal effectively with more than a small portion of the potentially relevant variables. With this recognition and with the help of Fisherian design principles, educational experiments have increasingly been conceived as multivariate investigations. Usually, several independent stimulus variables are manipulated and their joint effects on a single learning criterion are assessed. Such designs presume stimulus complexity but they ignore the possibility of response complexity. Experiments can also be designed to incorporate several different or repeated dependent response variables and/or to include, as additional independent variables, selections from a special class of antecedent response characteristics, here called "aptitudes". In classroom research, where students can be expected to bring widely different patterns of relevant prior experience to the experiment, and particularly in curriculum evaluation studies, where many different learning criteria might also be appropriately applied, experimental designs which are multivariate in this latter sense may be particularly important. The present paper considers the selection and use of response variables in the design of such investigations.

Four classes of response measurements can be distinguished. First, there are antecedent response variables, represented usually by scores on aptitude tests administered prior to an instructional treatment but including also sex, age, or any other index of a potentially important human difference. Second, the traditional conception of a single achievement test straddling an instructional treatment can sometimes be extended to permit repeated or intermediate measures at several points as instruction proceeds. This possibility can in turn be expanded to produce a third class of variables representing the many different learning effects to be assessed during instruction. Finally, there are the more remote or enduring effects of instruction as reflected in tests of retention, transfer and aptitudinal or attitudinal change.

These four kinds of measures are shown in Table I. Listed beneath each are some of the methods by which such variables have been or could be treated for the purposes of design and statistical analysis. The list is not exhaustive: it does not presume to survey in any general way the relevant statistical methods and design considerations already treated in detail by Tatsuoka and Tiedeman (1963) or by Campbell and Stanley (1963). It is intended rather to emphasize some points not made explicit in those two major sources and to publicize some more recent and ongoing developments. The table will not be described in detail but will be used instead to organize some more general comments about problems and possibilities for classroom research with respect to each class of variables.

Insert Table I about here

Aptitude Input Measures. Individual differences among students on aptitude variables have traditionally been viewed as a source of error to be controlled, in earlier days by matching procedures and, more recently , through the use of covariance analysis. A third view suggests that differential aptitudes should often be systematically included in experiments rather than being covaried out of them. Shreds of evidence so far available from the many studies that have incidentally correlated aptitude variables with learning under different instructional methods, or from the few investigations aimed specifically in this direction, indicate that the possibility of aptitude-treatment interactions deserves serious consideration. The demonstration of intersecting regression lines for the two treatments, that is, a disordinal interaction, implies that one instructional treatment is best for one group of students while another treatment is best for a different group. Such findings have practical as well as theoretical significance. They provide decision rules for the assignment of students to different paths toward the same instructional objective and they provide insights into the nature of aptitude functioning. For a fuller discussion of the importance of disordinal interactions, see Cronbach (1957) and Cronbach and Gleser (1965).

TABLE I

	Independent Response Variables	Dependent Response Variables		
	Variables	Occasions	Variables	Variables
STUDENTS	Aptitude Input Measures	Repeated Learning Measures	Multiple Learning Measures	Aptitude Output Measures
INCOMPLETE CONTROL OR ANALYSIS	Matching	Simple gain Relative gain Residual gain (DuBois, 1962)	Arbitrary composite	Separate ANOVA
CONTROL	ANCOVA MANCOVA (Winer, 1962)	Sequential ANCOVA	Post-treatment ANCOVA (Gourlay, 1953) (Cox, 1958)	
ANALYSIS OF VARIANCE	Treatment X levels ANOVA (Lindquist, 1953) (Stanley, 1960) (Page, 1965)	Repeated measures ANOVA and Trend analysis (Gaito and Wiley, and Bock, in Harris, 1963) Factor analysis of variance (Gollob, 1966)	Hotelling's T^2 Dyadic ANOVA (Tukey, 1949) ANOVA with multiple dependent variables (Roy and Gnanadesikan, 1959) (Tukey, 1962) (Bock, in press) (Bock and Haggard, in press)	
REGRESSION ANALYSIS	Multiple regression analysis	Generalized learning curve components (Tucker, 1960) Cattell's Covariation Chart (Cattell, 1959) Three-mode factor analysis (Tucker, in Harris, 1963)	Multiple discriminant function (Cooley and Lohnes, 1962)	

Canonical correlation and factor analysis (Harris in Harris, 1963)
Inter-battery factor analysis (Kristoff, 1965)

It should be noted that main effects in an analysis of variance are meaningless in the presence of such interaction and variables not included in the experiment have no opportunity to demonstrate their interactive effects. In view of the large number of nonsignificant overall treatment comparisons obtained in instructional research, it is appropriate to ask how many studies mask such interactions by ignoring aptitudes and averaging over them. One can also ask how many reportedly significant main effects are misleading, for the same reason.

The manner in which aptitude variables can be used in experiments depends upon several considerations. If it is possible to assign individual students randomly to treatments and if one aptitude variable or several uncorrelated aptitude variables are to be used, then treatments X levels designs as discussed by Lindquist (1953) and Stanley (1960) are appropriate. If one must deal with intact groups, such as classes, then an approach suggested by Page (1965) is possible. The case of two or more correlated variables with either individual or group randomization has not been formally worked out, though with individual random assignment some useful approximations are possible. Various correlational approaches are also available.

The selection of likely aptitude variables, however, rests on an analysis of subtle differences existing between the instructional treatments under consideration. At present there are few guidelines, though one hypothesis suggests that the status variables used most frequently in the past (e.g. sex, age, and IQ) may not be the most useful for this purpose. Of greater value may be the more narrowly defined intellectual and personality characteristics and, perhaps, members of a newer class of individual variables referred to as cognitive styles and preferences. In some exemplary studies, aptitude measures have been developed specifically for the particular instructional comparisons of interest. It is quite possible that the most useful aptitudes for learning research may have no precedents in the literature of differential psychology.

Repeated Learning Measures. Laboratory experimentation on learning has traditionally used practice as the continuum along which knowledge or skill acquisition could be measured. Although "improvement with practice" or "change with repetition" are incomplete as definitions of learning, practice curves are thought to show important features of learning pheromena and are, therefore, commonly used in presenting data from the laboratory. In research on formal instruction, however, investigators have normally been unable to treat their data in this way. The pre vs. posttest comparisons or gains scores derived from these tests have had to suffice. Thus, the presence or absence of some degree of learning has been studied, but there have been few attempts to investigate the course of acquisition.

Acknowledging that more than two test administrations may be difficult to obtain, it is nonetheless true that educational researchers have not really considered the possibility. They have continued using the difference score despite its apparent faults. Consequently, the effects of repetitive achievement testing are not known. There have also been few attempts to build truly equivalent forms of aptitude or achievement tests. The problems of measuring change form the subject of a whole book recently edited by Harris (1963) and cannot be elaborated upon here. One observation that can be made, however, is that powerful methodology is available for the analysis of repeated measures data, if such data can be obtained. Even without a series of formally equivalent measures available, there is still the possibility of serial measurements of a rougher sort, perhaps using unit tests, and various kinds of correlational analyses.

It is unlikely that learning will be understood, or curricula adequately evaluated, by considering global instructional treatments as black boxes with achievement tests as measures of inputs and outputs. A more analytical view must be adopted which includes ideas about the sequencing, staging, or patterning of learning phenomena. Acquisition stage measures may then be used in combination with aptitude input measures to provide specific diagnostic clues for curriculum revision as well as a more basic understanding of the complexity of educational processes in general. Comparisons between instructional treatments on a unit by unit, or even item by item, basis clearly offer a finer grain analysis than global gain scores. In this connection, group comparisons using item sampling (Cronbach, 1963) and criterion referencing (Glaser, 1963) ideas deserve consideration.

Multiple Learning Measures. Part of the problem mentioned in the preceding section is that instructional research has been wedded for too long to the achievement test as the sole arbiter of theory and practice. The suggestion that other kinds of dependent variables might be used is not new, yet rarely are multivariate data actually collected. If collected, rarely are they analyzed in a way that capitalizes on the fact that several measures are available on the same subjects. The typical approach has been to treat each variable by a separate ANOVA and, at most, to examine a table of intercorrelations thereafter. Some more satisfying analytical methods do exist but they have not yet been fully developed or publicized. Concerted effort by statisticians in this critically important area has really only just begun. It is therefore not surprising that researchers have continued to think in terms of single dependent variables and separate analyses.

Nonetheless, it is increasingly apparent that the concern of curriculum research, for example, should be aimed at determining the effects of an instructional program, not merely its effectiveness (Cronbach, 1965). Acceptance of this revised goal creates a need for new methodology capable of handling a multidimensional conception of curriculum evaluation. In other areas of instructional research also, it is clear that the outcomes of instruction are potentially many and that a given stimulus variable may affect one kind of dependent variable and not another, or perhaps even affect the intercorrelation between them rather than the mean of either.

The most obvious instance of multiple criteria is the use of both response correctness and response latency. Another example might be the measurement of attitudinal or interest changes as well as cognitive changes due to instruction. But one can go a good deal further, especially if the multiple measures are also conceived as repeated measures. Some other possibilities are: various kinds of teacher or student ratings, quantified characteristics of student essays or extemporaneous written or oral classroom behavior, disciplinary or attendance of tardiness records, library and special resource usage, extent of various outside-of-school activities, etc., etc. An occasions X variables X students data cube such as that pictured in Table I could provide the basis for extensive analyses of teacher or curriculum effects and even studies in what might be called the ecology of classroom behavior.

Aptitude Output measures. Most of what was said about multiple learning measures is also true for aptitude output measures, so little more need be added about them. In fact, the distinction between the two classes, though deemed helpful, may be a bit arbitrary. The former class was defined to contain variables arising within an instructional treatment and readily adaptable to repeated measurement as well. The latter class refers to measures administered after the close of the treatment.

Presumably, long-term, relatively permanent effects of instruction are reflected in retention, transfer, and aptitude measures. Many instructional objectives, particularly those of the new curricula are in fact stated in transfer or aptitudinal terms (Cronbach, 1965). In some cases, new forms of instruction can be compared with older methods only in these terms, since no achievement test can be constructed without bias toward one or the other kind of content.

Until recently, such investigations used either separate ANOVAs or evaluated the retention or transfer effects after controlling immediate learning effects by covariance analysis. Although advocated by some authorities, this post-treatment covariance procedure cannot be recommended here. As in the case of multiple learning measures, the newer developments in multivariate ANOVA should supplant older, less complete modes of analysis. The future should also see increased use of methods for comparing aptitude batteries administered before and after instruction.

Summary. In summary, four emphases have been suggested as central concerns for either theoretically or practically oriented research. If instructional methods and processes are to be understood and improved, a much clearer conception is needed of:

1) the nature of student aptitudes as they interact with teaching and learning processes,
2) the course or patterning of these processes across the occasions on which they occur,
3) the extensity of instructional effects as well as the intensity of any one effect, and
4) the enduring changes and subsequent effects of learning relative to the pattern of intellectual development in general.

Research aimed at these goals must move from the upper half to the lower half of Table I for its methodology. This multivariate conception of learner input and learning outcome taxes our knowledge of traditional experimental design, while offering the hope that newer designs commensurate with the richness of classroom behavior will be increasingly available and increasingly applied.

Bock, R.D. Contributions of multivariate statistical methods to educational research. In Cattell, R. B. (Ed.) Handbook of multivariate experimental psychology. Chicago: Rand McNally, in press.

Bock, R.D. and Haggard, E.A. The use of multivariate analysis of variance in behavioral research. In Whitla, D. (Ed.) Handbook of measurement in education, psychology and sociology. Boston: Addison-Wesley, in press.

Campbell, D.T. and Stanley, J.C. Experimental and quasi-experimental designs for research on teaching. In Gage, N.L. (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 171-246.

Cattell, R. B. Personality. New York: McGraw-Hill, 1959.

Cooley, W.W. and Lohnes, P.R. Multivariate procedures for the behavioral sciences. New York: Wiley, 1962.

Cox, D.R. Planning of experiments. New York: Wiley, 1958.

Cronbach, L.J. The two disciplines of scientific psychology. Amer Psychologist, 1957, 12, 671-684.

Cronbach. L.J. Evaluation for course improvement. Teach Col R. 1963, 64, 672-683.

Cronbach, L.J. Issues current in educational psychology. Child Development Monogr. 1965, 30, 109-126.

Cronbach, L.J. The logic of experiments on discovery. Stanford University, 1965, 17p.

Cronbach, L.J. and Gleser, Goldine Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1965.

DuBois, P.H. The design of correlational studies in training. In Glaser, R. (Ed.) Training research and education. Pittsburgh: University of Pittsburgh Press, 1962. Pp. 63-86.

Glaser, R. Instructional technology and the measurement of learning outcomes: some questions. Amer. Psychologist, 1963, 17, 519-521.

Gollob, H.F. A statistical model which combines features of factor analysis and analysis of variance. Preprint 188, Mental Health Research Institute, University of Michigan, 1966.

Gourlay, N. Covariance analysis and its applications in psychological research. Brit J Stat Psychol. 1953, 6, 25-34.

Harris, C.W. (Ed.) Problems in measuring change. Madison: University of Wisconsin Press, 1963.

Kristoff, W. Orthogonal inter-battery factor analysis. RB65-29 Educational Testing Service, Princeton, New Jersey, 1965.

Lindquist, E.F. Design and analysis of experiments in psychology and education. Boston: Houghton Mifflin, 1953.

Page, E.B. Recapturing the richness within the classroom. Paper presented at American Educational Research Association, Annual Meeting, Chicago, Feb. 11, 1965.

Roy, S.N. and Gnanadesikan, R. Some contributions to ANOVA in one or more dimensions: I & II. Ann. Math. Stat. 1959, 30, 304-317, 318-340.

Stanley, J.C. Interaction of organisms with experimental variables as a key to the integration of organismic and variable-manipulating research. In 17th Yrbk natl Con. on Meas. in Educ., 1960, pp. 7-13.

Tatsuoka, M.M. and Tiedeman, D.V. Statistics as an aspect of scientific method in research on teaching. In Gage, N.L. (Ed.) Handbook of research on teaching. Chicago: Rand McNally, 1963. Pp. 142-170

Tucker, L.R. Determination of generalized learning curves by factor analysis. Educational Testing Service, Princeton, N.J. 1960.

Tukey, J.W. Dyadic anova, an analysis of variance for vectors. Human Biology., 1949, 21, 65-110.

Tukey, J.W. The future of data analysis. Ann. Math. Stat., 1962, 33, 1-67.

Winer, B.J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.